

## A KANJI INPUT SYSTEM WITH NONSTANDARD-CHARACTER PROCESSING

Masahiro MORIWAKI, Shigeo FURUKAWA, Yoshinari MITA, Mitsuhiro SAKAGUCHI, Nobuo NISHIDA, Tadafumi FUJIHARA and Fujio SAITO

Central Research Laboratories, Nippon Electric Co., Ltd.  
Kawasaki, Japan

A Kanji (Chinese ideograph) input system with nonstandard-character processing is described. About 2,000 characters, which are most frequently used in writing Japanese sentences, are encoded for identification. Moreover, a very many other characters, which may be often used in a particular text, are encoded in the same manner as those 2,000 characters after generation of their patterns and assignment of their codes were achieved interactively with a small hologram tablet and a monitor plasma display. A hologram is used as a code generator for the characters and the small tablet, and concentrates functions, such as character identification, Kanji pattern generation and their code assignments. The whole operation is achieved by using just a single stylus. In this paper, a Kanji pattern generation method is described. Descriptions of the system and its hardware are also given.

### 1. INTRODUCTION

Due to the recent remarkable growth of information processing industries in Japan, research and development on many kinds of input/output equipments, which can process Japanese language, especially the Kanji (Chinese ideographs) are in progress. A Kanji printer and a Kanji display, which are used as output equipments, may be realized comparatively easily with the application of usual techniques. However, a Kanji input equipment is not realized easily, because of the large number of Kanji, the several different ways of reading each Kanji and a need for human intervention. An equipment which makes it possible to encode all kinds of Kanji cannot be realized. As a result, in equipments in general use, 2,000 to 4,000 characters, which are most frequently used and can express almost all Japanese sentences, are treated as standard characters. Other Kanji are treated as nonstandard characters.

The Kanji Input System in this paper not only encodes standard characters, but also encodes nonstandard characters in the same manner as standard characters, after their codes and character patterns are entered into the system. As the character input section and the pattern input section are on the same plane, only one stylus can point out characters on the character input section, as well as draw character patterns on the pattern input sections. Character codes for characters and positional codes for character patterns are generated from only one holographic coding plate.<sup>(1)</sup> A plasma display panel is used as a monitor.

### 2. KANJI INPUT SYSTEMS PROBLEMS

Japanese characters consist of three types of character sets, kata-kana, hira-gana and the Kanji. All three types are used for orthography in Japan. The kata-kana and hira-gana are phonetic signs, both of which have only 50 characters in each set. However, the Kanji are ideographs and the set consists of more

than 50,000 characters. Therefore, the technical consideration for Kanji input systems are focused on how to enter so many different characters into computer memories in addresses which are readily accessible for comparison and readout, on an economical as well as efficient basis.

A Kanji input system in general use<sup>(2)(3)</sup> has only a standard character set. The characters of a given text are encoded by selecting them from the set. When the text characters are not in the set, they, the so-called nonstandard characters, are replaced with phonetic signs, e.g. hira-gana, or numbers of their predetermined codes, or special symbols for which explanations are given at the end of the text. They are not often used, but may be used very frequently in a certain part of the text during a certain period. (For example, "孝", the name of an ancient emperor, was used in the news about excavation of an old Chinese tomb.) In such a case, the above-mentioned character input methods occupy much operator time and are prone to cause many mistakes.

To help solve the above-mentioned Kanji input system problems, the Kanji Input System has been designed to embody the following features.

- (1) A nonstandard character pattern can be generated and entered into the system. The character pattern, once entered into the memory, can be used for an output equipment as well as for a monitor.
- (2) Nonstandard characters can be used in the same way as standard characters after their registration in the system. They can be entered into the system during a necessary period, and may be erased from the system after their use period is over.
- (3) Input mistakes in characters can be easily found and corrected by using the monitor, the plasma display panel.
- (4) As the character input section, the function input section and the pattern input section are on the same plane, operation of the system is achieved by using only one stylus.

(5) Character codes for characters and positional codes for character patterns are generated very simply from the holographic coding plate.

### 3. SYSTEM DESCRIPTION

Characters in the Kanji Input System are defined as follows: Characters are divided into two categories, standard and nonstandard characters. Standard characters are divided into regular and non-regular characters. Regular characters are those whose character codes and patterns are already defined in the system. Each of them is fixedly arranged on the reserved area of the character input section to allow direct encoding. Non-regular characters are those whose character codes and patterns are already defined in the system, but the area of the character input section set aside for them is not so fixed. Nonstandard characters are those whose character codes and patterns are not defined in the system. Non-regular and nonstandard characters are able to be encoded directly, when they are accommodated into the spare area of the character input section.

Figure 1 shows a block diagram of the system. In this figure, thick lines, thin lines, and dotted lines indicate processing flow, data flow and visual feedback to the monitor, respectively. Mode functions indicated by the thick line blocks are: (1) character pattern input for nonstandard characters, (2) accommodation of non-regular or nonstandard characters into the spare area of the character input section and (3) text input, each of which is carried out interactively with the monitor.

The input sections in the system consists of the character input section divided into the reserved

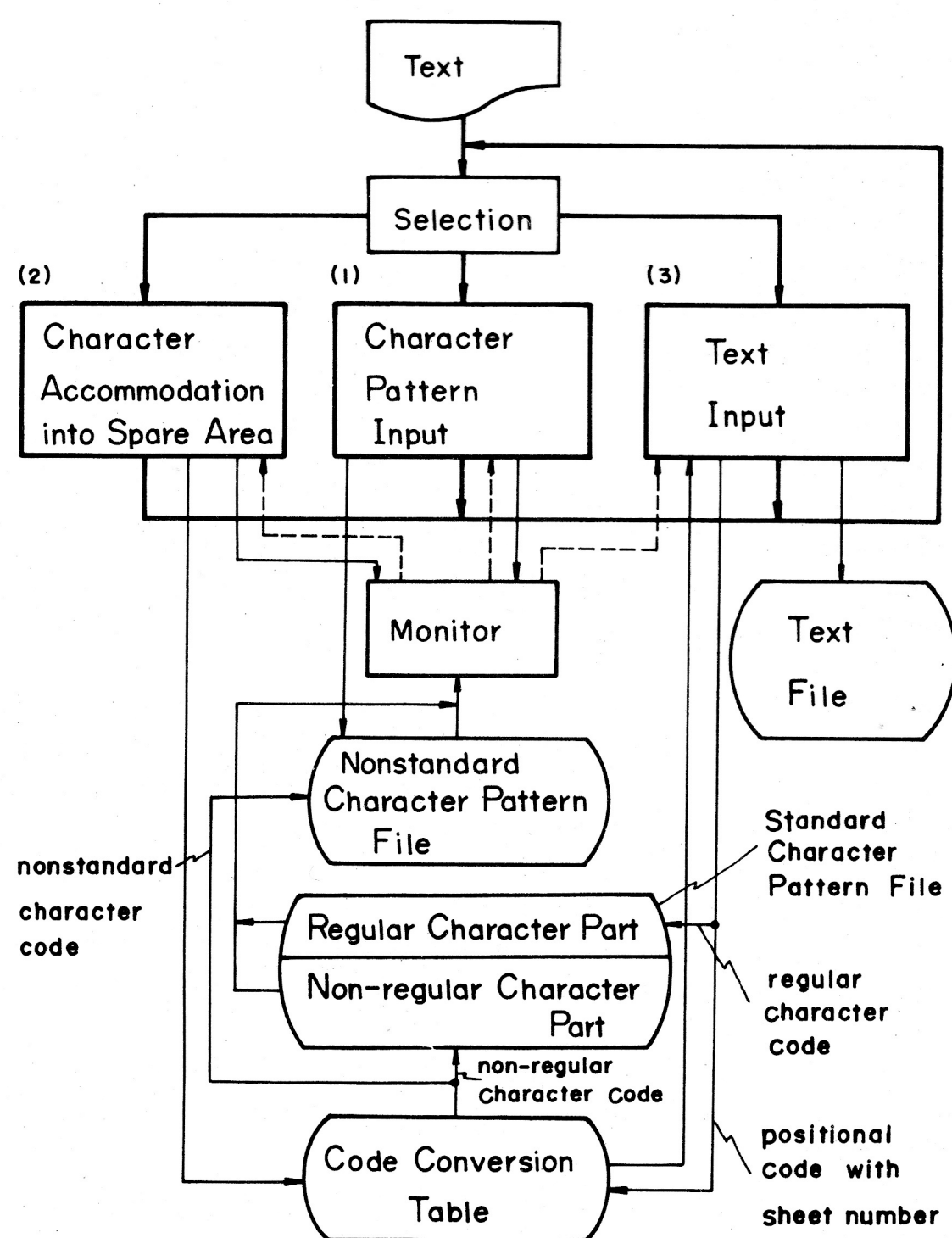


Fig. 1. System block diagram

area for regular characters and the spare area for non-regular or nonstandard characters, the pattern input section and the function input section.

The character pattern input mode is selected to generate nonstandard character patterns. Character pattern generation is carried out in the stroke type format, in which the beginning, turning and end points of each stroke are specified. The generated patterns are stored temporarily in the nonstandard character pattern file. An access to the newly stored pattern is possible by assigning a pertinent nonstandard character code.

The accommodation mode is selected to insert non-regular or nonstandard characters into the spare area. For nonstandard characters, the character pattern input mode is selected after or before the selection of this mode to accomplish character registration into the system. The spare area is occupied by an interchangeable sheet, on which accommodated character patterns are written by the operator. The accommodation procedure is the specification of the sheet number, character positions on the sheet (or positional code) and non-regular or nonstandard character codes. These specified results are stored in the code conversion table.

The text input mode is usually selected in order to encode the characters of the given text. The encoded text, consisting of regular, non-regular or nonstandard character code, is stored in the text file. The regular character is pointed out with the stylus on the reserved area. An accommodated non-regular character or a registered nonstandard character is pointed out on the sheet (the spare area). When the character exists on a different sheet from that currently in place, the lack is compensated for by inserting the sheet containing that character. Then the system is informed of the sheet number. The encoded character is displayed immediately on the monitor by accessing the standard or nonstandard character pattern file.

### 4. HARDWARE

The Kanji Input System uses a Kanji input equipment

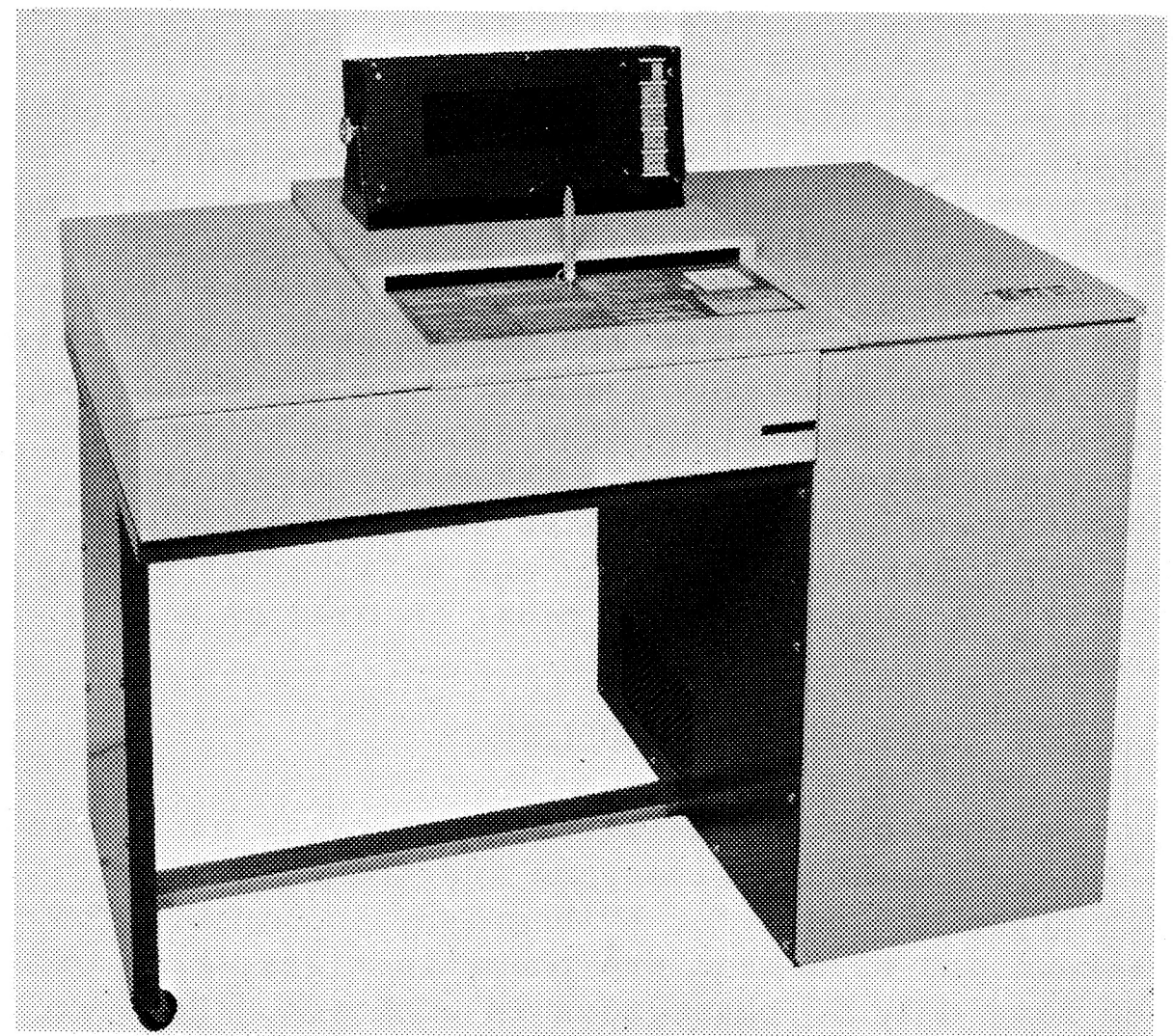


Fig. 2. Front view of the Kanji Input Equipment

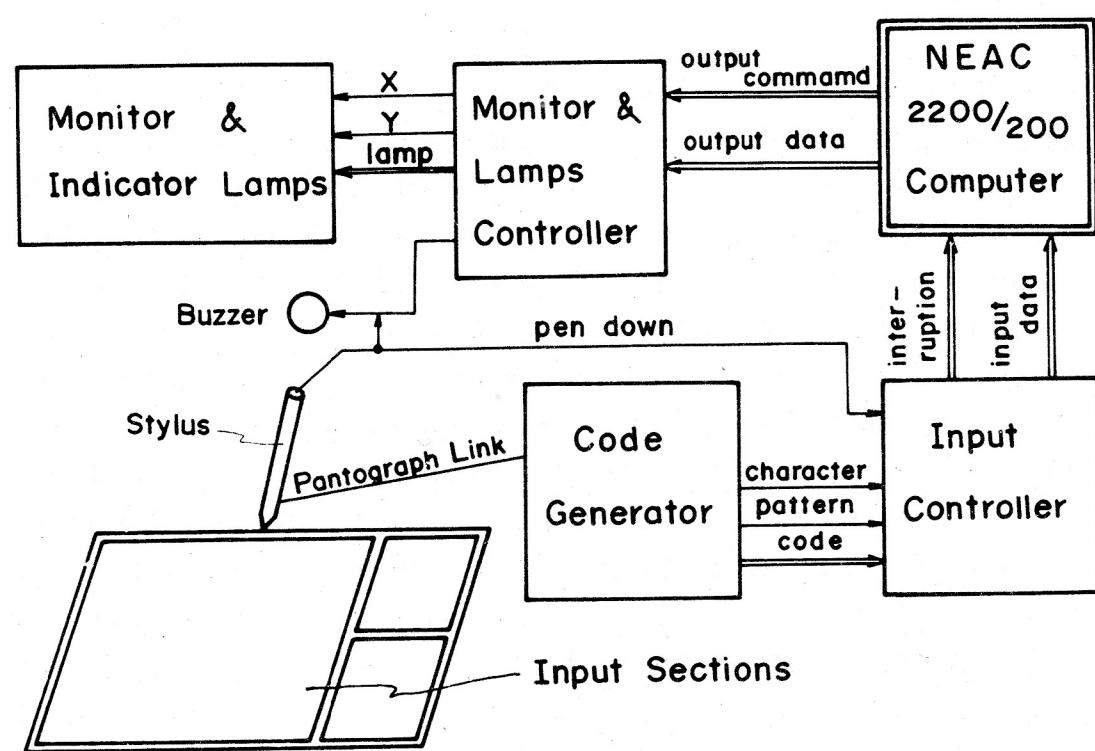


Fig. 3. Hardware configuration

(Fig. 2) and NEAC 2200/200 computer. The Kanji input equipment consists of input sections, a stylus, a code generator, a monitor and indicator lamps, as shown in Fig. 3.

#### 4.1 Input Sections

The input sections as shown in Fig. 4, have a character input section composed of a reserved area and a spare area, a function input section and a pattern input section. These sections are placed on the same plane. On the reserved area, 2,205 regular characters are arranged in phonetic order. These characters are composed of 82 alphanumerics, 75 hiragana, 75 kata-kana, 1,938 Kanji and 35 symbols. The character arrangement in this area is the same as the character board of the Kanji typing machine which is in general use in Japan. 36 non-regular and non-standard characters can be accommodated on the spare area. On the function input section, 45 kinds of functions can be assigned. The spare area and the function input section are on the same sheet, which is replaceable by an operator, as required, to change contents for various pertinent applications. The pattern input section is a small tablet, 93 mm  $\times$  75 mm in size, which has a resolution of 2 lines/mm.

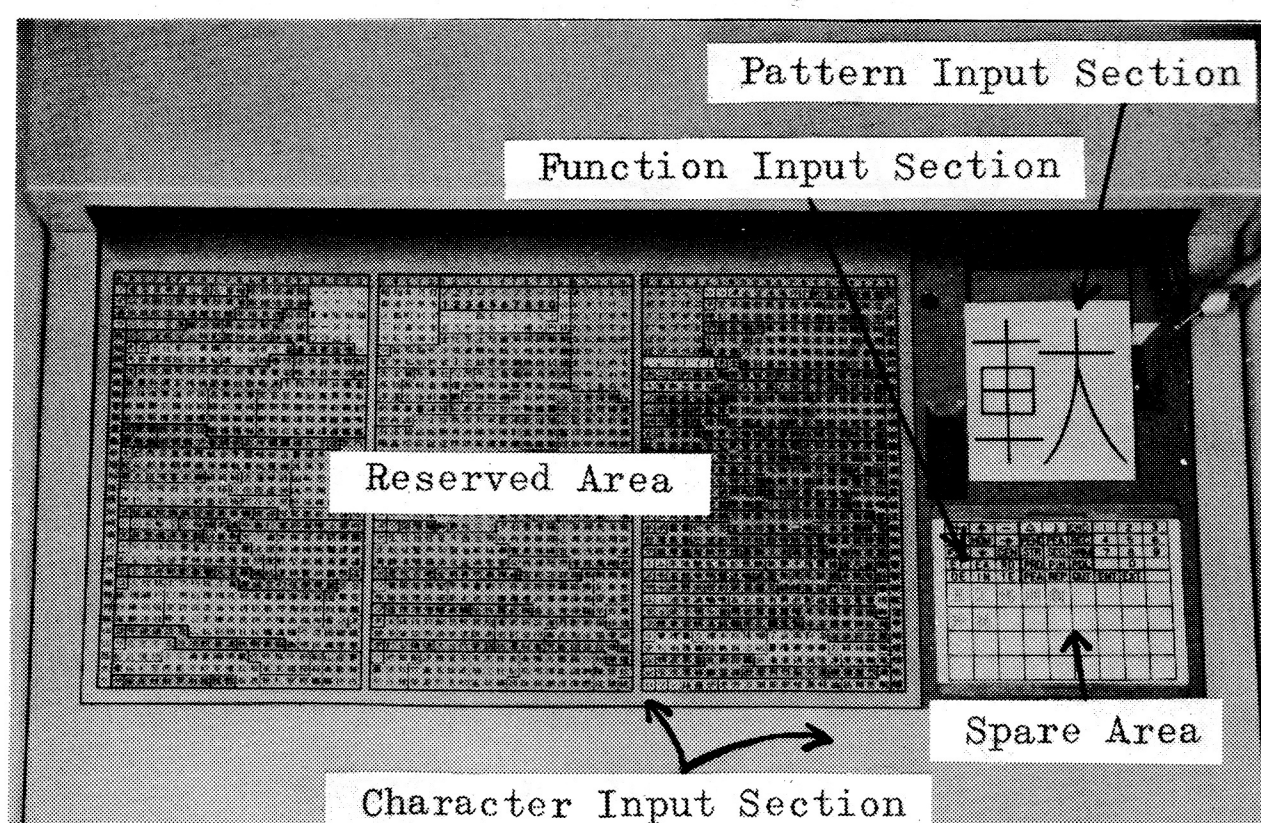


Fig. 4. Input sections

#### 4.2 Stylus

The stylus is utilized to point out desired characters on the character input section and functions on the function input section, as well as to draw patterns on the pattern input section. Input of the character codes and the function codes is achieved by putting the stylus into the conic hollows, which are located in the upper right position of the area occupied by the desired characters or functions. The stylus is connected with the pantograph link and its position is transferred to the code generator mechanically. The status of the switch in the stylus is an indication of the operator's action. Whenever the stylus is pressed, the "Pen down" signal, indicating stylus activation into the computer circuitry, is issued to the computer, and the operator is notified by the buzzer.

#### 4.3 Code Generator

The code generator makes use of a holographic coding plate to digitize effectively the position of the stylus on the input sections, which have various kinds of functions, as stated previously. As shown in Fig. 5, the pantograph link is connected at one end to the stylus and holds the light pen on its other end. The displacement of the light pen is about 1/4.4 of that of the stylus. The laser beam from the light pen reaches the holographic coding plate at a perpendicular angle of incidence, and is diffracted through it. The first order diffraction beams fall on the photocells, each of which is connected to an amplifier. As a result, electric codes are issued from the amplifiers, corresponding to the position of the stylus on the input sections.

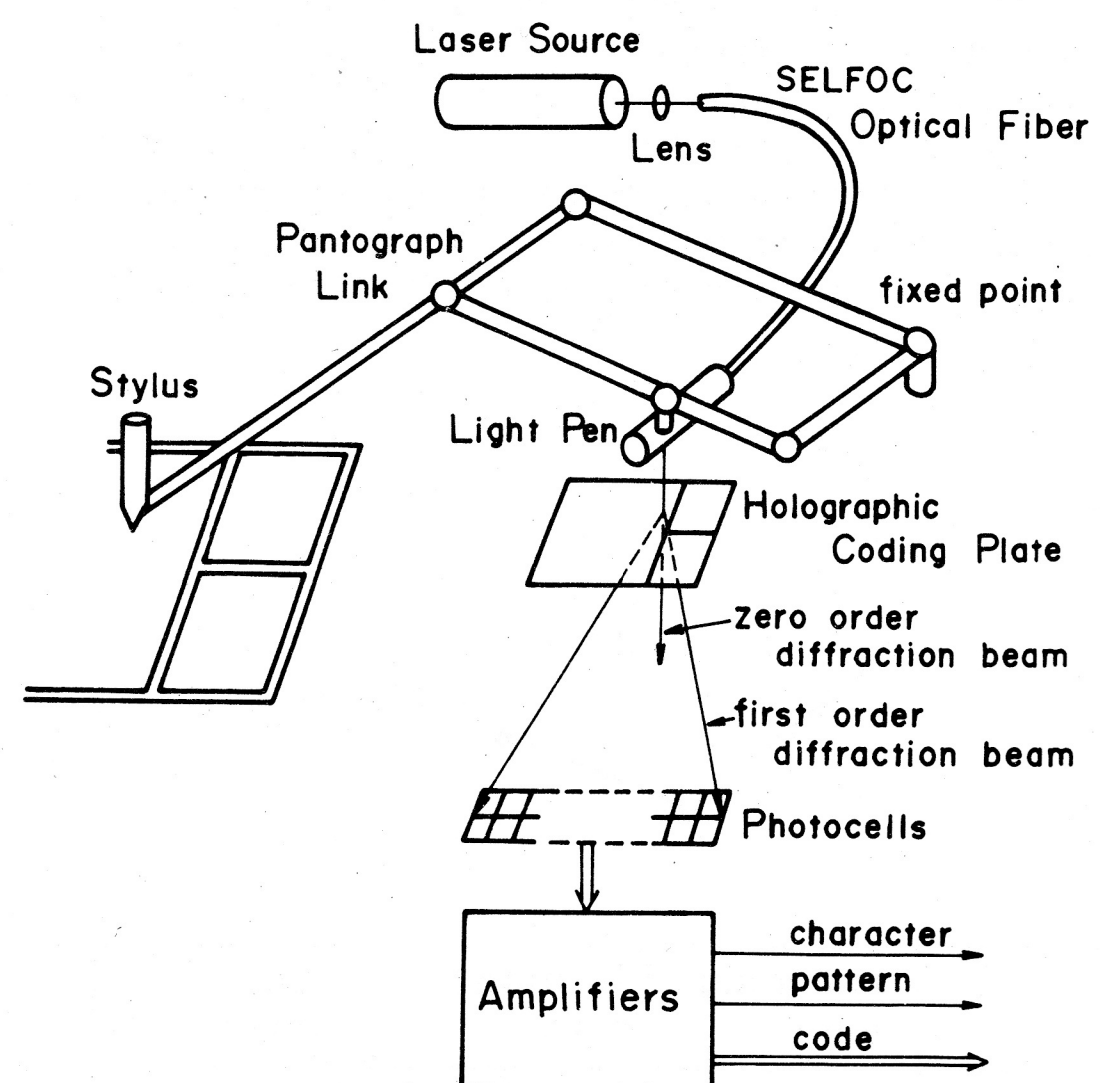


Fig. 5. Code generator

The code generator generates the 14-bit character code, a parity bit plus the "Character" bit which distinguishes the character code from the positional code when the stylus points out the character on the character input section or points out the function

on the function input section. Also, it generates the 16-bit positional code (X: 8 bits, Y: 8 bits) when the stylus draws the character pattern on the pattern input section.

#### 4.4 Monitor and Indicator Lamps

The monitor is a plasma display panel (PDP) which monitors the input character and the generated character pattern. The PDP is produced by Owens-Illinois Inc. and its individually addressable light points are  $256 \times 80$  dots. The PDP has a display capacity of 48 characters, each of which is composed of a  $16 \times 16$  dot matrix.

The indicator lamps indicate the present state of the system. The top indicator lamps indicate the input sections which are ready to use. The others show text input, pattern input, accommodation and error, respectively.

### 5. PROCEDURES FOR ENABLING ENCODING NON-REGULAR AND NONSTANDARD CHARACTERS

#### 5.1 Character Pattern Input for Nonstandard Characters

The character pattern input mode with the character accommodation mode for nonstandard characters and the character accommodation mode for non-regular characters are executed, respectively, to allow their direct encoding.

Character patterns are generated interactively in the stroke format by using the stylus and the monitor. A character consists of strokes. A stroke is composed of a number of straight lines called segments. The relation between a stroke and its segments is shown in Fig. 6. In this example, the stroke is composed of segments 1 and 2, and the stroke or the segments are specified by the beginning, turning (for stroke) and end points. The character pattern generated by  $64 \times 64$  grid points is reduced to  $1/4$  in size and then added to the repertoire in the nonstandard character pattern file. The pattern generation consists of three phases: stroke generation, stroke correction and segment correction.

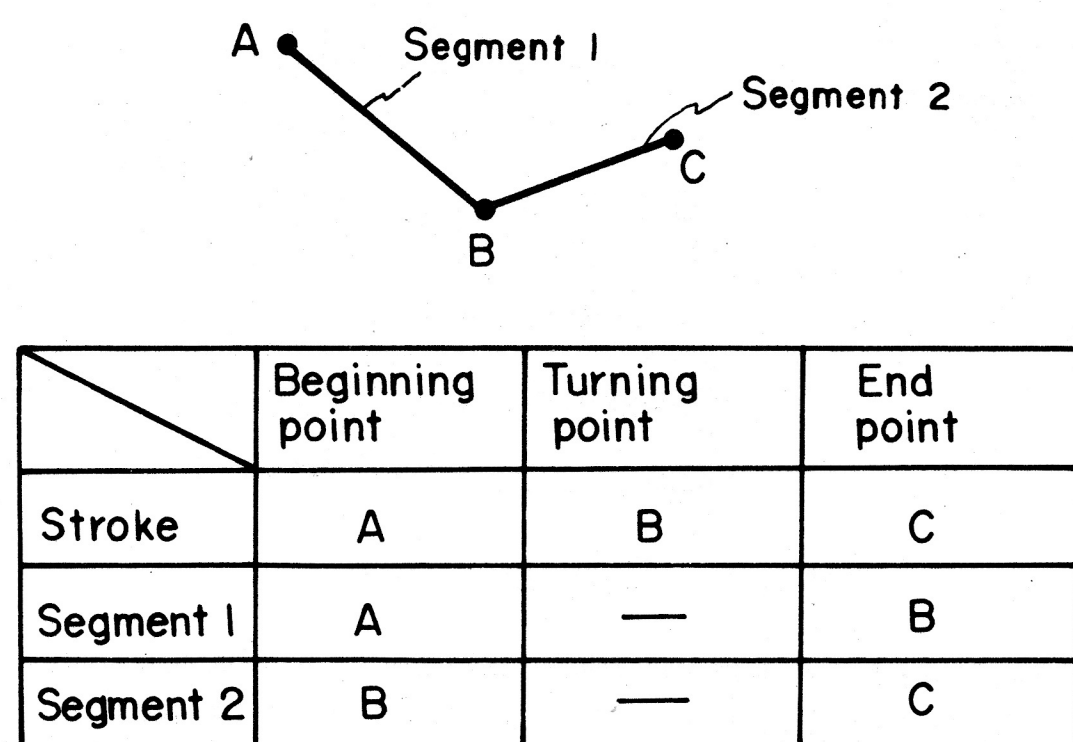


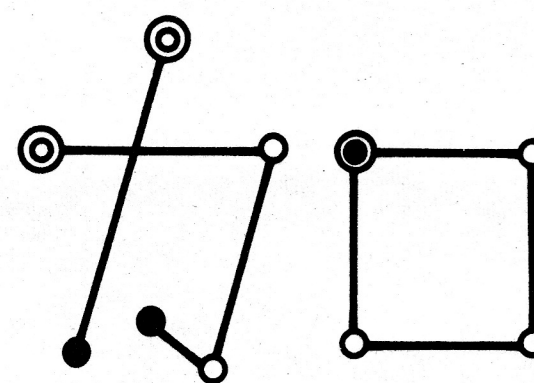
Fig. 6. Relations between a stroke and its segment

#### (1) Stroke generation phase

Strokes are generated in this phase. First, the rectangle indicating the work domain, consisting of  $64 \times 64$  grid points, is displayed on the monitor. The operator sets the beginning point of a stroke by touching the stylus to the equivalent point on the previously drawn character pattern on the pattern input section. The stylus movement is checked by the cursor (blinking point) movement in the work domain. The specified beginning point is steadily lit on the monitor. For setting the next turning or end point of the stroke, the stylus is lifted and moved and then touched to the equivalent point on the drawn pattern. By the action, the vector connecting the beginning point with the newly set point is drawn in the work domain. If a stroke is composed of several segments, the stylus is moved and touched to the region near the last turning point for the next segment input. By this specification, the beginning point of the next segment is generated internally so that this point is coincident to the end point of the preceding segment. Character pattern "力口" is generated by specifying the beginning, turning and end points of strokes, as shown in Fig. 7.

#### (2) Stroke correction phase

The stroke to be corrected is selected by touching the stylus to the region near the beginning, turning or end point of that stroke on the pattern input section. Then, the selected stroke blinks. If there are several strokes whose beginning, turning or end points are the same or very nearly so, the earliest generated one is selected. By specifications of "Before" or "After" functions, a stroke generated before or after the currently selected stroke is selected. The selected stroke is moved in the work domain related to the stylus movement from the touched point on the pattern input section. This movement stops when the stylus is lifted from that section. The "Delete" function deletes the currently selected stroke.



- ◎ Beginning point of a stroke
- Turning point of a stroke
- End point of a stroke
- ◎ Beginning and end point of a stroke

Fig. 7. Stroke generation example

### (3) Segment correction phase

This phase can be called only from the stroke correction phase. The first segment of the selected stroke in the stroke correction phase is automatically selected and blinks. If another segment is desired, the function "After" or "Before" is used, just like in the stroke selection. The segment is corrected by moving the end point related to the stylus movement from the touched point on the pattern input section. This movement of the end point also means movement of the beginning point of the next segment. The specification of the "Delete" function causes the deletion of the end point of the currently selected segment. Figure 8 shows the generated character pattern "車大" on the monitor, where the large Kanji on the right is the manipulated result after corrections in the work domain and the small Kanji on the left is the reduced pattern.

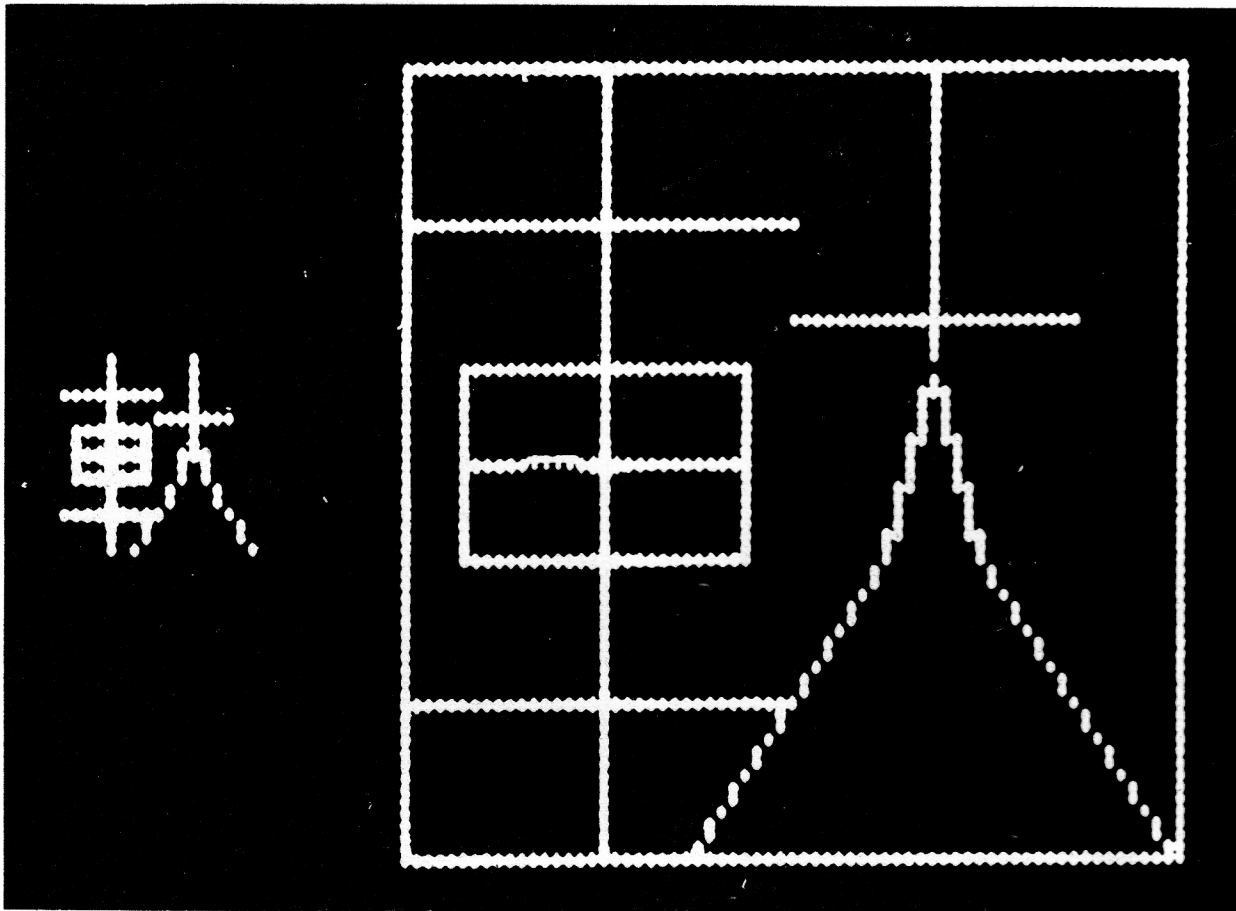


Fig. 8. Generated character pattern example

### 5.2 Accommodation of Non-regular and Nonstandard Characters in the Spare Area

The accommodation procedure includes the specification of the sheet number, the character position on the sheet and the code of non-regular or nonstandard character. The previously assigned sheet number is specified by using the "Digit" (0~9) function. The character position or equivalent positional code is specified by touching the stylus to the desired one of 36 positions on the sheet. The character pattern for the accommodated character is drawn by the operator on that specified position of the sheet to enable easy access. The code of a non-regular or nonstandard character is assigned by using the "Digit" function. The non-regular character code is a predetermined code. However, the nonstandard character code is newly assigned by the operator and that code forms a different group from that of the standard character code. The specified result is stored in the code conversion table.

### 6. CONCLUSION

The Kanji Input System described in this paper can treat literally an infinite number of characters

with a small capacity character pattern file, since it generates patterns and codes of nonstandard characters for itself. It solves the problem of the number of characters in a Kanji input system being limited because of its difficulty to operate, although a large number of characters in Japanese language is often required for actual sentences. The system has been given a trial in the management system of technical reports in NEC's central research laboratories.

### ACKNOWLEDGEMENT

The authors wish to thank Mr. S. Ohkura for his helpful advice in preparing this paper and Mr. K. Kiji and Dr. T. Uchida for their continued guidance. The authors also thank Mr. M. Shutoh and Mr. K. Naitoh for their help in the installation of the PDP, Mr. Y. Yamakawa, Mr. K. Nakazawa and Mr. T. Terada for their designing the system, and Mr. K. Kosuge for his manufacturing the holographic coding plate.

### REFERENCES

- [1] M. Sakaguchi and N. Nishida, The Hologram Tablet — A New Graphic Input Device, *AFIPS Conf. Proc. FJCC*, Vol. 32, pp. 653-658, 1970.
- [2] M. Moriwaki, M. Sakaguchi, N. Nishida, T. Takada and I. Hirabayashi, Kanji Input Device, *Bulletin of 1972 Annual Meeting of Inst. Electronics Comm. Engrs. Japan*, P.1340 (Japanese), April 1972.
- [3] M. Moriwaki, M. Sakaguchi, N. Nishida, T. Fujihara, F. Saito, Y. Mita and Y. Kikuchi, A Kanji Input Device Using a Hologram Coding Plate, *Bulletin of the 13-th Annual Meeting of Information Processing Society of Japan*, P.329 (Japanese), December 1972.
- [4] F. Saito, Y. Mita, M. Sakaguchi, N. Nishida, M. Moriwaki, T. Fujihara, I. Hirabayashi and T. Takada, *Holotablet*, NEC R & D, No. 29, pp. 1-9, April 1973.